Searching for hidden descriptors in the metal-ligand bond through statistical analysis of DFT results

Oier Lakuntza,[†] Maria Besora,[†] and Feliu Maseras[*,†,‡]

†*Institute of Chemical Research of Catalonia (ICIQ), The Barcelona Institute of Science and Technology, Avgda. Països Catalans, 16, 43007 Tarragona, Catalonia, Spain*

‡*Department de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain*

E-mail: fmaseras@iciq.es

Phone: +34 977920202. Fax: +34 977920231

Abstract

A statistical treatment of the DFT-computed heterolytic bond dissociation energies (BDE) between a diverse variety of metal fragments and ligands leads to the identification of five hidden descriptors that best characterize the bonding ability per moiety, and of a simple mathematical formula able to obtain from these hidden descriptors a BDE estimation within a few kcal/mol from the DFT value. A simple extension of this treatment beyond the original set of metal fragments and ligands is also presented. The first two hidden descriptors can be associated with the well-known concepts of σ-donation and π-effects, with the next two associated to *cis* influence and degree of covalency. The procedure can be easily extended to additional ligands and metal fragments, and opens the way to an improved understanding of fundamental concepts of chemical bonding.

Introduction

Bond Dissociation Energy (BDE) is a fundamental concept in chemistry. It measures the strength of the interaction between two fragments and it gives key information on their relative stabilities and potential reactivity. BDEs are particularly useful in the case of transition metal complexes, as a lot of the reactivity of these species operates through ligand exchange processes, with thermodynamics ruled by differences in BDE. The thoroughly used orbital analysis of the metal-ligand bond aims to a qualitative prediction of the BDE. Its application to transition metal chemistry has allowed the identification as main contribution to bonding of the σ interaction between a doubly occupied valence orbital of a ligand and a empty valence d orbital of the transition metal.[1,2] Orbital analysis has also led to the identification of potential π-interactions, such as the back donation from the metal to a π-acceptor ligand.[3–6]

An elegant option for the translation of qualitative concepts such as σ donor or π acceptor strength to a quantitative description of BDE is the introduction of descriptors. These are sets of numbers associated to a ligand or a metal fragment that can be introduced in a mathematical formula to yield a quantitative value for the desired property, in this case the BDE. The concept of descriptor is at the base of the quantitative structure-activity relationships (QSAR) tools of widespread use in chemical biology. The σ values and related parameters used in the Hammett equation and other free-energy relationships of widespread use in physical organic chemistry are also descriptors. Similar free-energy relationships between BDEs have been tried in organometallic chemistry.[7–12] Although the use of descriptors is less widespread in transition metal chemistry, a number of significant contributions have been made. They are often focused on a particular set of complexes, or on a particular reaction. Among the descriptors focused on a ligand type, one must mention the pioneering work by Tolman on the design of steric and electronic parameters for phosphorus donor ligands,[13] which have been later refined by other authors.[14–18] Descriptors have been also developed for other families of ligands such as carbenes,[19–24] and C-donors,[25] and for the

quantification of the π-donating and back-donating ability of chemical fragments.[26–28] An alternative approach has been the design of descriptors for specific reactions or reaction types, often with the goal of improving catalyst design. One must mention in this concern the work by the group of Bo and co-workers on 3D molecular descriptors to predict catalytic enantioselectivity,[29–31] and more recently Carbó and co-workers to describe the nucleophilicity of trivalent boron compounds.[32] Other descriptors aimed at quantifying the sterics are the Accessible Molecular Surface (AMS)[33] and related stereocartography tools.[34] Cavallo and co-workers proposed an analysis of the surface of interaction catalyst/substrate to quantify steric and electrostatic effects.[35] Jensen has applied descriptors to the automated design of organometallic molecules from fragments derived from crystal structures or calculations.[36–40] Worthy contributions on predictive modeling have been made by the groups of Rothenberg[41–44] and Paton.[45,46] These reaction-oriented descriptors have also been used in organic chemistry by the group of Sigman.[47–52]

In most of the cases described above, descriptors are obtained from molecular properties, ranging from simple properties as orbital energies or cone angles to elaborate constructions as steric estimations from the curvature of the molecule surface. A remarkable twist is introduced in the work by Fey and co-workers,[17,18,53,54] where starting from molecular descriptors, statistical procedures based in principal component analysis are carried out to define new descriptors and generate a ligand knowledge base. Similar concepts have been also applied by Qu and co-workers on the prediction of BDE energies inside organic molecules.[55]

A complementary approach is supplied by machine learning and neural network techniques that have started to be applied with success in the field[56] for instance to predict energies,[57–59] the electronic structure,[60] reaction yields,[61] or molecular atomization energies.[62]

In the current work, we intend to go one step further in the statistical treatment. Instead of using pre-defined descriptors, we are going to apply statistics to compute the best possible descriptors, which we will refer to as hidden descriptors. We will use computed heterolytic

BDEs as the starting point of the analysis. We are aware that the IUPAC definition of BDE involves homolytic cleavage. But we consider that the heterolytic cleavage of a metal-ligand bond is more interesting in the fields of inorganic and organometallic chemistry as the ligand exchange processes are governed by these energies rather than those of the homolytic process. Thus, unless otherwise stated, we will consider the heterolytic process $(\text{M-L})^x \rightarrow \text{M}^{(x-y)} + \text{L}^y$ were no radical species are being formed.

These hidden descriptors are non-conventional and should recover well-known factors such as donation and back-donation, and may discover new ones. In order to carry out a significant statistical treatment we need a large number of data. Luckily, modern advances in computational power and quantum chemical methods allow for efficient calculations of large number of BDEs. Two other factors favoring the application of statistics to this problem are the existence of a natural heterolytic cleavage partition facilitating the definition of BDE, and the availability of large number of potential complex-ligand couplings. We have carried out BDE calculations on all the combinations between 42 metal fragments and 43 ligands, for a total of 1806 BDE calculations. As we are interested in learning about the fundamental aspects of the bond between metal and ligand, we tried to use fragments as diverse as possible. The only caveat was to reduce as much as possible obvious steric effects, as we consider that these are easier to characterize. For the sake of simplicity, a preliminary analysis is carried out on BDEs in vacuum, although a subsequent analysis of values in water solvent is found to be more productive.

This article is organized in a series of sections that follow our construction of the model. This introduction is followed by brief descriptions on the procedure to construct the **BDE** table and the statistical procedure to obtain the hidden descriptors. A first analysis on vacuum BDEs is presented, and found to be insufficient in terms of a qualitative analysis of the hidden descriptors. Then an analysis on BDEs in water solution is presented, and found to be more satisfactory. The last sections describe a general procedure to compute new hidden descriptors for metal fragments and ligands outside of the training set, and the

modifications that take place in the analysis when dispersion effects are removed.

Results and discussion

Organization of the results

For the obtention of the hidden descriptors and the corresponding estimated BDEs we followed the scheme depicted in Figure 1. We build first a table with the heterolytic BDEs of different metal fragments and ligands ($(\text{M-L})^x \rightarrow \text{M}^{(x-y)} + \text{L}^y$). We perform a statistical analysis consisting of a matrix decomposition. We reduce the dimensionality of the matrices obtained from the decomposition. This process results in a set of hidden descriptors for metal fragments and also for ligands. Using them we can reconstruct a new BDE matrix with the estimated BDEs, which can then be compared with the initial one.

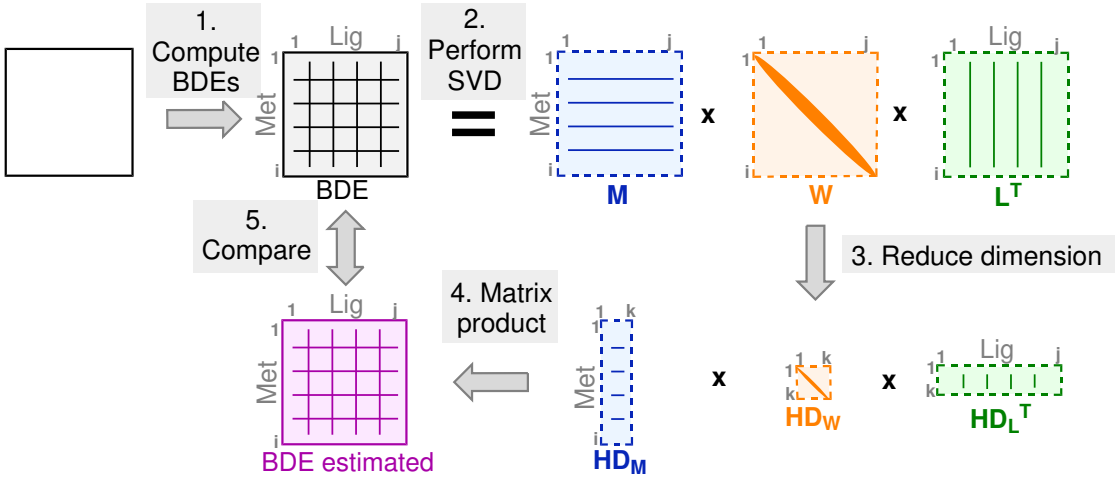


Figure 1: Schematic presentation of the procedure.

Building of the BDE table

For the initial set of calculations (initial training set), we used a series of 22 ligands and 23 metal fragments. The 22 ligands were Br^-, CH_3^-, Cl^-, CN^-, CO, F^- $C_6H_5^-$, H_2, H_2O, H^-, I^-, NCH, NCS^-, NH_3, NHC, O^{2-}, OH^-, PH_3, C_5H_5N, S^{2-}, SCN^- and SH^- (see Figure 2). The idea was to choose a variety of the most common ligands. This included ligands with

charges 0 (CO, PH_3, NH_3), -1 (Cl^-, H^-, OH^-) and -2 (O^{2-}, S^{2-}). Pure σ-donors such as H^- and strong π-acceptors as CO were included. A variety of coordinating atoms were considered. As we did not want to focus on steric effects, bulky substituents were avoided. The 23 metal fragments were $AlCl_3$, $AuPH_3^+$, $Co(NH_3)_5^{3+}$, CrO_3, $Cu(NH_3)_3^{2+}$, $FeCl_2^+$, $Fe(CO)_4$, $IrCO(PH_3)_2^+$, K^+, MnO_3^+, $Mo(SH)_3^+$, $Nb(NH_2)_4^+$, OsO_3^{2+}, $PdH(PH_3)_2^+$, $PdPH_3$, PtF_5^-, $Rh(H_2O)_5^{3+}$, $Ru(SH)_4$, $TaMe_4^+$, $TiCl_3^+$, $W(CO)_5$, $ZnCH_3^+$ and $ZrCl_5^-$. The criteria to select metal fragments was to choose representative complexes for a variety of metals and delete one of their ligands. For instance, $Co(NH_3)_5^{3+}$ was obtained from $Co(NH_3)_6^{3+}$ and OsO_3^{2+} from OsO_4. We included transition metals from the three periods, with inclusion also of early and late transition metals. We added also $AlCl_3$ and K^+, without transition metal, because we considered them representative examples of dative bonding. We carried out geometry optimizations of each of these metal fragments and ligands, and of their 506 combinations. All geometry optimizations and energy calculations were carried out with the Gaussian 09[63] suite of programs. Optimizations were performed with the hybrid generalized B3LYP-D3[64,65] functional where empirical dispersion was added using Grimme's dispersion correction.[66] We used the 6-31+G(d)[67,68] basis set for elements between H and Cl, and Stuttgart/Dresden effective core potentials (ECP) together with the corresponding SDD basis sets for heavier atoms. Harmonic vibrational frequencies were calculated to ensure that the structures found are true minima. All energies reported correspond to potential energies in gas phase plus zero-point energy corrections. A data set collection of computational results is available in the ioChem-BD repository.[69]

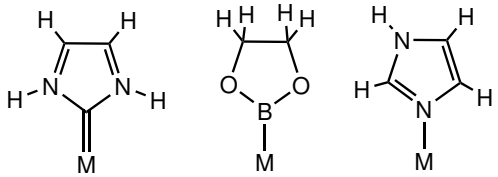


Figure 2: Representation of some of the ligands where the formula supplied in the text could be ambigous.

We applied this procedure to the generation of data for extra sets of 21 new ligands and 19 new metal fragments. This new group of ligands and complexes will be used to check the validity of the conclusions extracted from the initial training set. The new ligands were: $C_2H_4O_2B^-$ ($Bpin^-$), CCH^-, $CHCH_2^-$, $C_6H_4Cl^-$, $C_6H_4NO_2^-$, $C_6H_4OMe^-$, H_2S, He, $C_3H_4N_2$ (imidazole), NH_2^-, NMe_3, C_4H_4O (THF), OMe^-, PCl_3, PF_3, SiH_3^-, $SiMe_3^-$, SMe^-, C_4H_4S (thiophene) and Xe. The new metal fragments were: $Ag(NHC)^+$, $AuCl_3$, AuCN, $Cr(H_2O)_5^{2+}$, $CuCH_3$, CuCN, $FeCl_3$, $FeCl_3^-$, $GaCl_3$, HgI_2, $InCl_2^+$, $Ni(PF_3)_3$, $NiPH_3$, $PtPH_3$, $SnMe_3^+$, $Y(H_2O)_5^{3+}$, $ZnCl_3^-$, $ZnNH_3^{2+}$ and $Zr(OMe)_3^+$. With these new ligands and complexes we have a total of 43 ligands and 42 complexes, generating a table of 1806 heterolytic BDEs.

Each metal fragment, ligand and resulting complex were computed at their ground state. In a few cases the ground state of the resulting complex depends on the interacting ligand or the ground state of the metal fragment is not equal to the ground state of the complex. In these cases the most common ground state was used for all ligands and for the metal fragment in order to avoid variations related to change in spin multiplicity. The isomeric nature of each complex and ligand was also preserved; each ligand was coordinated in the same mode to all complexes. Details of the multiplicity and geometry of each complex can be found in the Supporting Information.

We wrote some in-house scripts, included in the Supporting Information, to run with the statistics package Octave[70] and automatically perform the Singular Value Decomposition (SVD), matrix operations and the regressions to compute estimated interaction energies and their errors. We finally performed single point calculations on the B3LYP-D3 optimized structures without the Grimme's dispersion correction (B3LYP) and in water through the use of polarizable continuum model (PCM) to build additional Tables that will be discussed in later sections.[71,72]

Statistical analysis: matrix decomposition

The procedure outlined in the previous section produced a 23x22 matrix **BDE** with all bond dissociation energies. We will describe in this section our procedure to extract information from this matrix in general terms, for a more accurate description please see the Supporting Information.

This procedure starts with a singular value decomposition analysis (SVD).[73,74] SVD is a statistical procedure, that decomposes a matrix into three new matrices. In this work, the $\mathbf{BDE}_{i,j}$ between a metal fragment i and a ligand j can be written as in equation 1. **M** is an i x i unitary matrix related to the metal fragment, **W** is a i x j diagonal matrix known as the singular values of **BDE** and **L** is a j x j unitary matrix related to the ligands ($\mathbf{L}^T$ is the transposed of the **L** matrix). The contribution of ligands and metal fragments is separated by this analysis.

$$\mathbf{BDE}_{i,j} = \mathbf{M}_i \cdot \mathbf{W} \cdot \mathbf{L}_j^T \tag{1}$$

The key feature of the SVD is that the importance of the weights **W** is hierarchical with the first one being the most important. So, SVD can be used to reduce a high-dimensional dataset into fewer dimensions while retaining the most important information. This is done by reducing the number of terms from j to a smaller number k. The $\mathbf{M}_i$ matrix is pruned from a i x i matrix to i x k matrix that we will call $\mathbf{HD}_M$. The $\mathbf{L}_j$ matrix will be reduced from j x j to j x k and will be called $\mathbf{HD}_L$ and similarly the diagonal matrix **W** from i x j to k x k and will be $\mathbf{HD}_W$. As the weights are ordered by their importance taking the first k terms results in taking the most relevant information. With that, the recovered $\mathbf{BDE}_{i,j}$ matrix will have the original i x j size but will be less accurate (of lower rank). The lack of accuracy comes from the dimension reduction in matrices **M**, **W** and **L**, that will be smaller, see Figure 1. Part of the information that was originally contained in BDE will not be recovered (from k to j and from k to i) and will be a source of error.

Application of the matrix decomposition to our data in vacuum

A good reason to apply SVD to a set of data is that it allows to reduce the dimensionality of the data. The decomposition outlined in equation 1 for matrix **BDE** is exact as far as it uses 22 singular values. As detailed above a pruned version of equation 1, will be able to capture the main features of the full matrix.

We carried out a preliminary statistical analysis to evaluate the representativity of our initial training set and we replaced three of the ligands and three of the complexes by elements in the initial extended set. The details of this tuning of the training set are detailed in the Supporting Information. The actual training set includes: the 22 ligands Br^-, CH_3^-, Cl^-, CN^-, CO, F^-, $C_6H_5^-$, H_2, H_2O, H^-, NCH, NH_3, NHC, O^{2-}, OH^-, PH_3, C_5H_5N, S^{2-}, SH^-, $C_6H_4OMe^-$, PCl_3 and $SiMe_3^-$ and the 23 metal fragments $AuPH_3^+$, $Co(NH_3)_5^{3+}$, CrO_3, $Cu(NH_3)_3^{2+}$, $FeCl_2^+$, $Fe(CO)_4$, $IrCO(PH_3)_2^+$, MnO_3^+, $Mo(SH)_3^+$, $Nb(NH_2)_4^+$, OsO_3^{2+}, $PdH(PH_3)_2^+$, $PdPH_3$, PtF_5^-, $Rh(H_2O)_5^{3+}$, $Ru(SH)_4$, $TaMe_4^+$, $TiCl_3^+$, $W(CO)_5$, $ZrCl_5^-$ $AuCl_3$, HgI_2 and $InCl_2^+$. We will use the actual training set as training set for the rest of the study.

The number of required hidden descriptors (k) to consider for each metal and ligand depends on the particular problem at hand, namely the internal correlation among the data and the expected accuracy. Because of this we applied versions of equation 1 with different numbers of hidden descriptors and evaluated their performance. The quality of the performance was measured through comparison between the estimated BDE and the exact one. Both maximum error and average unsigned error were evaluated. Selected results are collected in Table 1 and Figure . The mathematical foundation of the procedure leads to an expected reduction of the error as one increases the number of hidden descriptors, and this is confirmed in the Table. But the reduction is not homogeneous. It is sharp at the beginning, and it slows down after a few hidden descriptors are included. With six terms, the maximum and average errors are 8.1 and 1.4 kcal/mol, respectively. The BDE energies that have these associated errors range between +39 to -912 kcal/mol. These values seem acceptable to us,

especially taking into account the large absolute values involved in gas phase BDEs (+54 − -912 kcal/mol).Moreover, the improvement associated to further expansion of the number of hidden descriptors is modest. Because of this, we settled in six terms, the version of equation 1 we are going to use in the remainder of the article will use a diagonal matrix $\mathbf{HD}_W$ with six elements, and six hidden descriptors for each metal fragment $\mathbf{HD}_M$ and each ligand $\mathbf{HD}_L$.

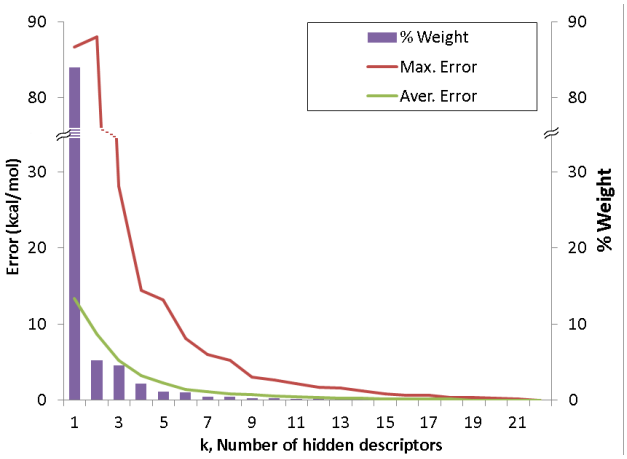


Figure 3: Performance of the hidden descriptors for the BDE values in vacuum as a function of the number of terms considered. Maximum error (red line), average error (green line) and relative weight (purple bars).

The treatment outlined above reduces the electronic properties in vacuum of each ligand and metal fragment to six hidden descriptors. Each ligand and metal fragment is characterized by a vector with six components, which are given in the Supporting Information. For instance, Br^- is characterized by the (HD_L= 0.190, 0.217, 0.301, 0.096, -0.124, -0.123) vector, and $AuPH_3^+$ is characterized by the (HD_M= -0.178, 0.091, -0.008, 0.167, 0.099, 0.186) vector. The introduction of these vectors in equation 1 will provide a reasonable approach to the BDE in any ligand-metal fragment combination. Each element of these vectors is a descriptor, and its main role is to provide a quantitative tool for the evaluation of BDEs. It would be interesting, however, to see how these descriptors can be associated to our intuitive concepts about bonding.

Let us clarify first that the sign of the **M** and **L** matrices, which give rise to the descriptors, is arbitrary. It is obvious that if we change the sign in all numbers in both the **M** and the **L**

Table 1: Maximum and average error (in kcal/mol) of the estimated BDE in vacuum with respect to the number of hidden descriptor terms considered.

	Actual Training set			
k	Max. Error	Aver. Error	Weight $\mathbf{W}_{k,k}$	% Weight
1	86.7	13.4	4691.4	84.0
2	88.0	8.7	292.7	5.2
3	28.1	5.2	252.3	4.5
4	14.4	3.2	119.3	2.1
5	13.2	2.3	59.0	1.1
6	8.1	1.4	54.1	1.0
7	6.0	1.1	24.6	0.4
8	5.2	0.8	22.0	0.4
9	3.0	0.7	15.5	0.3
10	2.6	0.5	12.6	0.2
12	1.7	0.3	7.4	0.1
14	1.2	0.2	4.4	0.1
16	0.6	0.1	2.5	0.0
18	0.3	0.1	1.3	0.0
20	0.2	0.0	0.9	0.0
22	0.0	0.0	0.5	0.0

matrix the results will be the same, as the matrix $\mathbf{W}$ contains only positive values. However, the signs between the two matrices are interconnected. A ligand with a positive sign for a given descriptor will have an associated binding contribution with metal fragments where the descriptor has a negative sign, and a repulsive one with metal fragments where the descriptor has a positive sign.

The association between are hidden descriptors and intuitive concepts is to a certain point subjective, and we are going to use as a guide the correlation between the values of our hidden descriptors and a large set of conventional descriptors, 666, for the ligands. Most of the 666 conventional descriptors are derived from applying computational chemistry tools to the ligands and some L−M complexes, others have been obtained from tabulated values. Among the considered conventional descriptors there are different types of charges, energies of the orbitals, values extracted from Bader, NBO, NRT and CDA analyses, geometry related descriptors such as different types of volumes, M-L bond distances, buried volume,... and different ways of computing chemical concepts such as ionization potential, electron affinity,

chemical potential, hardness,... among others. The complete list of conventional descriptors used can be found in the Supporting Information. We have searched, through an Octave script, the correlation between each of the six hidden descriptors ($HD_{L(1-6)}$) and each of the 666 conventional descriptors. All data corresponding to the application of this criterion are provided in the Supporting Information, and only a few selected values will be discussed here.

The first hidden descriptor in vacuum, HD_{L1}, shows a high correlation with a number of conventional descriptors. It shows a high correlation with the total charge of the ligand with an r^2 value of 0.946. It correlates even better with the energy of the HOMO orbital (r^2= 0.956). Anionic ligands are stronger donors, have higher total charges and have HOMO orbitals with higher energies. This hidden descriptor can be thus easily associated to donation, electron transfer, from the ligand to the metal.

The finding of σ donation playing a key role in metal-ligand bonding is encouraging. Unfortunately, results for other hidden descriptors in vacuum are less satisfying. The second hidden descriptor has the highest correlation (0.830) with a molecular descriptor corresponding to the APT charge of the ligand in the L$-FeCO_4$ complex. APT charges correspond to the atomic polar tensor derived charges, that it is one of the many ways to estimating partial charges. This seems associated to electrostatic correlations after the electron transfer has taken place. The best correlation of the HD_{L3} hidden descriptor with the list of conventional descriptors is r^2= 0.840 for the square of the Mulliken charge on the central atom divided by the volume of the ligand. Mulliken population analysis charges are another way of estimating partial charges. In this case it is the charge divided by the volume, meaning that it is not the overall charge what is important, but the "concentration" of charge in the ligand. This may be associated to softness/hardness, although the connection is not straightforward, as explicit descriptors for these magnitudes were also present in the set. The fourth hidden descriptor HD_{L4} shows a good correlation, r^2= 0.775, with the subtraction of ionic and covalent interaction on the L-$TaMe_4$ complex according to the Natural Bond Order analysis.

The Natural Bond Order analysis is one of the available tools of the Natural Bond Order calculations (NBO). It divides the bond order of a given bond in two terms: ionic and covalent. The subtraction of both terms gives and idea about which is the most relevant. This seems associated to the covalency of the chemical bond. The chemical meaning of the last two hidden descriptors is even less clear, as the best correlations between the 666 conventional descriptors and HD_{L5} and HD_{L6} are only 0.656 and 0.586.

We were particularly puzzled by the fact that there was no obvious connection between the hidden descriptors and the standard descriptors related to π-bonding. π interactions are generally accepted to play an important role in metal-ligand bond, and their absence from our set of hidden descriptors was conspicuous. We tried to evaluate the performance of our analysis in other circumstances. In particular, the results with B3LYP without dispersion were very similar to those reported above, and are shown in the Supporting Information. In contrast, the effects of solvation, discussed below, are substantial.

Application of the matrix decomposition to our data in water

The unusual nature of some of the descriptors obtained from the table of the BDEs in vacuum prompted us to reevaluate the validity of our treatment. After all, the chemistry of inorganic complexes is almost always carried out in solution. It might well be that our intuitive concepts on chemical bonding are related to reactions in solution, and they do not apply well to reactions in gas phase. So we decided to repeat the analysis reported above, but now for species in water solution. In order to do that, we recomputed, through single point calculations, all the bond dissociation energies, and repeated the analysis outlined above. We chose water as a representative solvent for transition metal chemistry. The fact that it has a high dielectric constant, very different from vacuum, was also appealing. Calculations performed model the solvent through the use of a continuum model, explicit water molecules are not considered, the dielectric constant of the solvent and the shape of the solvent/solute is taken into account. In this approach hydrogen bonds between the solvent and the solute

cannot be taken into account.

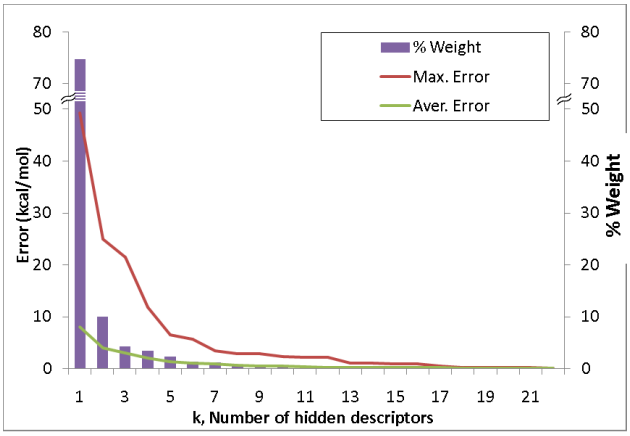


Figure 4: Performance of the hidden descriptors for the BDE values in water as a function of the number of terms considered. Maximum error (red line), average error (green line) and relative weight (purple bars).

Table 2 and Figure 4 present the associated error to each number of hidden descriptors for the BDEs in water. This must be compared to Table 1 and Figure for BDEs in vacuum. The structure of the Tables is similar, but there are some significant differences. A first one is that the number of required hidden descriptors terms drops from six to five. Using 6 terms the maximum and average errors are of 5.6 and 1.0 kcal/mol, using 5 hidden descriptors 6.5 and 1.3 kcal/mol respectively. The BDE energies that have these associated errors range between +20 to -280 kcal/mol. There is an overall very good agreement between the computed BDEs and the estimated ones. Figure 5 shows computed BDEs against the estimated ones using five hidden descriptors and the correlation is very good, r^2= 0.998. The importance of each hidden descriptor, that is the weight $\mathbf{W}_{k,k}$ changes also significantly. The weight associated to the first descriptor is 1529, while that of the second descriptor is 204. These values in water contrast with those of 4691 and 293 in vacuum, respectively. While in vacuum the value for the first weight $\mathbf{W}_{1,1}$ is sixteen times larger than that for the second $\mathbf{W}_{2,2}$, the ratio is reduced to seven times larger in water. Obviously, water stabilizes better separated charges. Not only the relative weight of the hidden descriptors changes, but also their nature, which can be checked in the full list provided in the Supporting Information. The first hidden

Table 2: Maximum and average error (in kcal/mol) of the estimated BDE in water with respect to the number of hidden descriptor terms considered.

	Actual Training set in water			
k	Max. Error	Aver. Error	Weight $\mathbf{W}_{k,k}$	% Weight
1	49.3	8.0	1529	74.7
2	24.9	4.0	204	10.0
3	21.5	2.9	85	4.2
4	11.8	2.0	69	3.4
5	6.5	1.3	47	2.3
6	5.6	1.0	25	1.2
7	3.4	0.8	22	1.1
8	2.8	0.6	12	0.6
9	2.8	0.5	10	0.5
10	2.2	0.4	9	0.4
12	2.1	0.2	6	0.3
14	1.0	0.2	4	0.2
16	0.9	0.1	3	0.1
18	0.2	0.0	2	0.1
20	0.1	0.0	1	0.0
22	0.0	0.0	0	0.0

descriptor in water correlates $r^2 = 0.81$ with the first hidden descriptor in gas phase, from the second onwards the hidden descriptors do not fully agree in vacuum and water.

We analyzed the correlation of each of the five hidden descriptors identified for BDE in water with the 666 standard descriptors discussed above for vacuum. The response was in this case much closer to expectation than for vacuum. Correlation tables are supplied in the Supporting Information, and key values discussed below. The ligands and metal fragments having the largest and smallest values for each descriptor are shown in Tables 3 and 4.

For the first hidden descriptor, a number of standard descriptors present high correlations. We find significant the correlation of 0.893 with the difference between the total charge of the free ligand and the total Mulliken charge of the ligand in the L−$Fe(CO)_4$) complex, see Figure 6 . This corresponds literally to the transfer from the ligand to the metal fragment in this particular set of complexes. This is associated to donation, mostly σ donation. This was also the case for the first hidden descriptor in vacuum, and corresponds to expectations. If we look at the values of this hidden descriptor for each ligand in Table 3 we see that

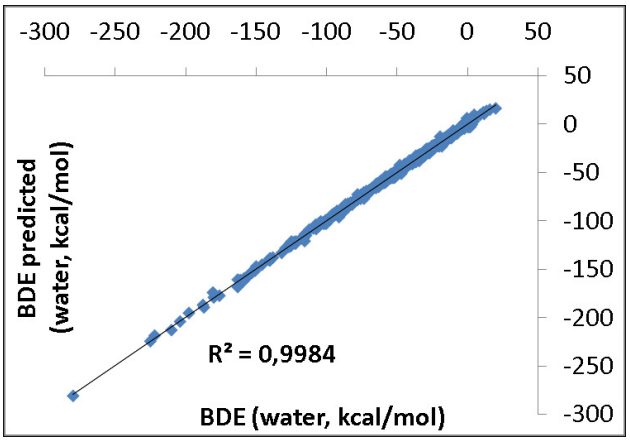


Figure 5: Plot of computed BDEs *vs* estimated BDEs for the actual training set in water.

Table 3: Ligands having the largest and smallest hidden descriptor value for each $\mathbf{HD}_L$ in water.

HD_{L1}	HD_{L2}	HD_{L3}	HD_{L4}	HD_{L5}
O^{2-} 0.42	O^{2-} 0.63	S^{2-} 0.43	CO 0.37	F^- 0.33
$SiMe_3^-$ 0.32	S^{2-} 0.25	Br^- 0.39	F^- 0.34	CH_3^- 0.21
H^- 0.32	OH^- 0.17	SH^- 0.34	NCH 0.27	OH^- 0.21
CH_3^- 0.31	F^- 0.16	Cl^- 0.33	H_2O 0.24	H_2O 0.18
...	...	...	...	...
H_2O 0.06	H^- -0.22	CH_3^- -0.10	$C_6H_5^-$ -0.16	PH_3 -0.26
PCl_3 0.05	PCl_3 -0.25	CO -0.29	H^- -0.18	CO -0.36
CO 0.04	$SiMe_3^-$ -0.26	O^{2-} -0.31	$C_6H_4OMe^-$ -0.20	O^{2-} -0.38
H_2 -0.01	CO -0.30	H^- -0.40	$SiMe_3^-$ -0.40	PCl_3 -0.53

almost all values are positive, ranging from 0.42 to -0.01. All the ligands selected donate to the metal fragments, the highest donation corresponding to the O^{2-} oxo, and the smallest to the dihydrogen H_2. The placement of the ligands corresponds highly to expectations. Examination of the values for the metal fragments in Table 4 is somehow less intuitive, but it can be rationalized. The weaker σ-acceptors contain d^{10} transition metals such as Hg^{2+} and Pd^0, and the strongest σ-acceptors contain d^0 metals in high oxidation state such as Os^{+8} and Mn^{+7}.

The second hidden descriptor presents the highest correlation (0.884) with a descriptor related to π bonding from the NBO analysis of $L{-}TiCl_3^+$ complexes. Specifically, the second order perturbation E(2) associated to the π orbitals in this complex. This analysis takes all

Table 4: Metal fragments having the largest and smallest hidden descriptor values for each $\mathbf{HD}_M$ in water.

HD_{M1}	HD_{M2}	HD_{M3}	HD_{M4}	HD_{M5}
HgI_2 -0.08	$Fe(CO)_4$ 0.31	$Nb(NH_2)_4^+$ 0.39	HgI_2 0.70	$PdPH_3$ 0.39
$PdPH_3$ -0.10	PtF_5^- 0.28	$ZrCl_5^-$ 0.35	OsO_3^{2+} 0.37	$Mo(SH)_3^+$ 0.34
$Nb(NH_2)_4^+$ -0.11	$AuCl_3$ 0.27	$Ru(SH)_4$ 0.26	$FeCl_2^+$ 0.16	OsO_3^{2+} 0.30
$ZrCl_5^-$ -0.11	$W(CO)_5$ 0.22	HgI_2 0.24	$AuCl_3$ 0.15	MnO_3^+ 0.16
...	...	...	...	...
$InCl_2^+$ -0.27	$Nb(NH_2)_4^+$ -0.25	$Rh(H_2O)_5^{3+}$ -0.13	$PdPH_3$ -0.21	$TaMe_4^+$ -0.19
$Rh(H_2O)_5^{3+}$ -0.29	$ZrCl_5^-$ -0.26	$InCl_2^+$ -0.18	$InCl_2^+$ -0.22	$Cu(NH_3)_3^{2+}$ -0.34
MnO_3^+ -0.35	$TiCl_3^+$ -0.36	MnO_3^+ -0.25	$TiCl_3^+$ -0.23	$InCl_2^+$ -0.39
OsO_3^{2+} -0.42	$TaMe_4^+$ -0.37	OsO_3^{2+} -0.50	CrO_3 -0.25	$Co(NH_3)_5^{3+}$ -0.47

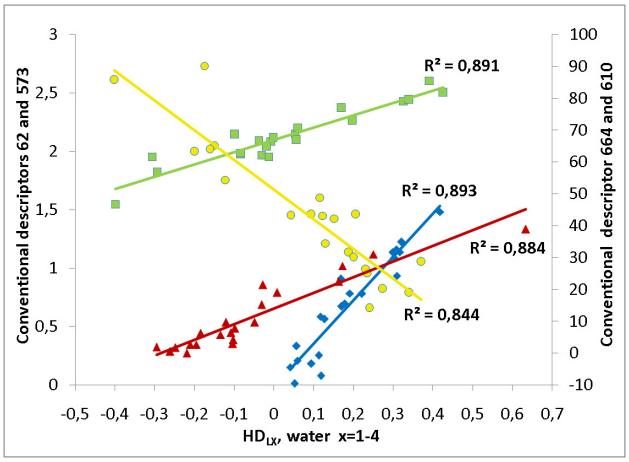


Figure 6: Plot of the values of the four first ligand hidden descriptors in water *vs* the values for the corresponding conventional descriptor giving the best correlation. $\mathbf{HD}_{L1}$ in blue, $\mathbf{HD}_{L2}$ in red, $\mathbf{HD}_{L3}$ in green and $\mathbf{HD}_{L4}$ in yellow.

interactions between Lewis-type NBOs (donor) and non-Lewis NBOs (acceptor), and gives their energetic contribution according to the second-order perturbation theory. This strongly suggests an association of the second hidden descriptor to π interactions, which is further confirmed by inspection of Tables 3 and 4. Here the values are both positive and negative. π-donating ligands have a favorable interaction with π-accepting metal fragments, but an unfavorable one with π-donating metal fragments. The strongest π-donating ligands are the dianionic O^{2-} and S^{2-} and the strongest π-accepting ligands are CO and $SiMe_3^-$. It is surprising to find H^- with a value of -0.22, thus π-accepting. This is obviously a misfit with the literal description of π-acceptor, as this ligand has no π orbitals. We rationalize it by

pointing out that the strong σ donor ability of hydride may increase the energy of occupied metal π orbitals, which are not hindered by ligand orbitals in this region.

The interpretation of the third hidden description is less obvious in first analysis. It correlates well (0.891) with a geometrical descriptor consisting on the distance between iron and the central atom of the ligand in the L−$Fe(CO)_4$ complexes. Initially, we found this puzzling, as the size of the central atom is not directly related to electronic properties. This hidden descriptor does not seem related either to steric effects, as CH_3^- appears in the small size region, and PCl_3, $SiMe_3^-$ have rather neutral values. Remarkably, the connection between atomic radius and this third hidden descriptor is lost when analyzing the values for the metal fragments in Table 4. Second row transition metals Nb, Zr appear in one end; while third (Os) and first (Mn) row transition metals appear at the opposite end. We are able to propose a qualitative explanation for this descriptor when realizing that metal fragments on one end have L−M−L close to 90° corresponding to trigonal bipyramidal or octahedral geometries, while on the other end they have values closer to 109°, corresponding to tetrahedral arrangements. We assign thus this third hidden descriptor to a *cis* influence. The lone pairs from ligands getting closer to the metal center will have a certain repulsion with the lone pairs of other ligands. We are aware that *trans* influence is much more common than *cis* influence, but we consider that *trans* influence is already included in σ donation and π interactions, thus in the first two hidden descriptors.

The fourth hidden descriptor correlates quite well (0.844) with an NBO descriptor from the L−$TiCl_3^+$ family of complexes. This is the percentage of covalency in the L-Ti bond. As mentioned above the Natural Bond Order analysis gives the ionic and covalent contribution to the bond order. This hidden descriptor also correlates with other covalency/ionicity descriptors for this and other complexes. We consider thus safe to assign this fourth descriptor to the covalency of the bond. According to Table 3, the ligands which favor covalent interactions are $SiMe_3^-$, H^- and the phenyl derivatives. A Lewis structure with these ligands containing a single bond with the metal would be quite reasonable. In the other end of the

Table, such a Lewis structure is less reasonable with ligands such as CO and F^-.

We could not find any good correlation between the fifth hidden descriptor and the standard descriptors. The highest value was 0.521, far below those reported above for the other hidden descriptors. This means that this fifth descriptor either corresponds to a property which is not reproduced by any of our 666 standard descriptors, or that it is ruled by the fitting to a numerical statistic noise. We will continue research in this issue, but we consider this is not critical for the current publication.

The analysis of the correlation between our hidden descriptors and an extended set of standard ones has allowed us to give a qualitative interpretation to their meaning. It is important to notice however that these standard descriptors cannot replace those computed by us in terms of performance, as typical r^2 values are in the range of 0.90, and they do not provide thus full reproduction of results.

Extension to other ligands and metal fragments

The procedure above provides hidden descriptors for the 22 ligands and 23 metal fragments in the actual training set. We discuss here the procedure to obtain new hidden descriptors for ligands and complexes outside the initial set. The key concept is to carry out a minimum number of BDE calculations for each new ligand or metal fragment, and use the correlations in the initial BDE matrix to estimate the corresponding hidden descriptor values, and hence the corresponding estimated BDEs in a procedure outlined in detail in the Supporting Information. A general presentation of the procedure is outlined in Figure 7.

We present here the results for BDEs in vacuum as a numerical example. For each new ligand, the BDE is computed with each of the following representative metal fragments: $AuPH_3^+$, $Cu(NH_3)_3^{2+}$, OsO_3^{2+}, PtF_5^-, $Ru(SH)_4$ and $TiCl_3^+$. From the six calculations for a given ligand, a set of hidden descriptors for this ligand can be obtained by assuming that the same linear relationships between BDEs present in the initial set are conserved. A similar treatment is carried out for new metal fragments. In this case, the BDE is computed

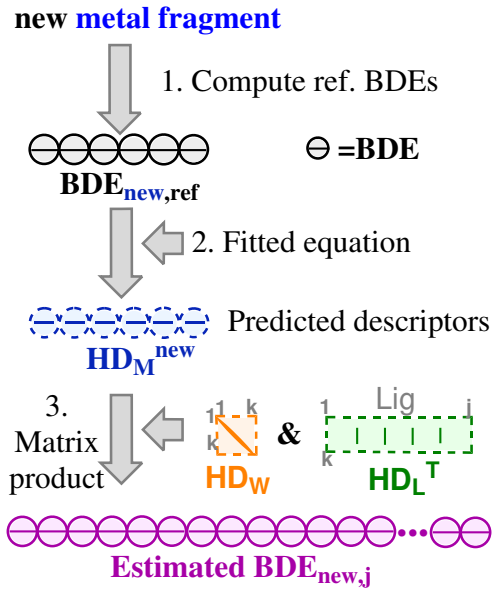


Figure 7: Procedure followed for the calculation of hidden descriptors for new metal fragment outside of the actual training set.

with each of the following representative ligands: Cl^-, H_2O, H^-, O^{2-}, $C_6H_4OMe^-$ and PCl_3. We applied this procedure to the generation of hidden descriptors for extra sets of 21 new ligands and 19 new metal fragments (the total minus those in the actual training set). We then computed the BDE for each metal fragment/ligand pair and compared the response with those provided by equation 1. The average absolute error for both metal fragment and ligand in the initial set was 1.4 kcal/mol, and the maximum error was 8.1 kcal/mol as mentioned above. The average and maximum absolute errors for a ligand in the initial set and a new metal fragment were 2.0 and 16.5 kcal/mol, and that for a new ligand and a metal fragment in the initial set were 2.3 and 12.9 kcal/mol. The BDE energies that have these associated errors range between +39 to -912 kcal/mol for the training set, between +63 to -783 kcal/mol for new metal fragment, between +45 to -525 kcal/mol for new ligands and between +53 to -407 kcal/mol for a new ligand and a new metal fragment. Finally, the average and maximum absolute errors for both metal fragment and ligand in the extra sets was 3.0 and 11.6 kcal/mol. The errors are obviously larger when extra fragments are taken into account, but we consider the increase acceptable. The correlation between the

computed BDEs and the estimated ones is r^2=0.99985 for the training set, r^2= 0.9994 for new metal fragments with ligands in the training set, r^2= 0.9992 for new ligands and metal complexes in the training set and r^2= 0.998 for new metal complexes and new ligands, see Figure 8. In fact, these results confirm that our training sets of ligands and metal fragments are sufficiently diverse to reproduce the behavior of external moieties.

The responses obtained by the methodology described in this work are very promising and open a broad area of application. We expect to expand this type of analysis to more ligands and metal fragments. We intend also to to perform in the nerar future this type of analysis for new ligands and metal fragments in water.

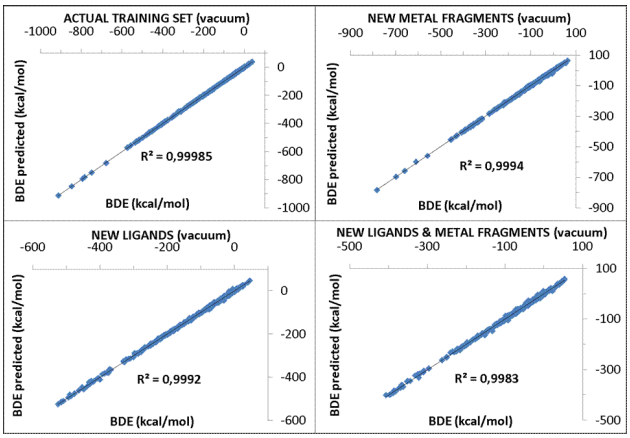


Figure 8: Plot of computed BDEs *vs* estimated BDEs in gas phase. Actual training set (top left), ligands in the training set and new metal fragments (top right), new ligands and metal fragments in the training set(bottom left), new ligands and new metal complexes (bottom right).

Conclusion

We have developed a statistical procedure leading to the identification of a limited set of hidden descriptors able to characterize the heterolytic bond dissociation energies (BDE) between arbitrary sets of ligands and metal fragments ($(M\text{-}L)^x \rightarrow M^{(x-y)} + L^y$). Application of this treatment, based in Singular Value Decomposition (SVD), to the computed bond dissociation energies (BDE) in water between a diverse variety of 22 ligands and 23 metal

fragments allows the identification of five hidden descriptors per fragment. Introduction of these hidden descriptors in a simple mathematical formula leads to estimated BDEs with an average error of 1.4 kcal/mol from those of the DFT calculation. We have also developed a procedure to obtain hidden descriptors for ligands and metal fragments outside of the initial set, a procedure that requires only a few additional calculations per fragment. This scheme has been applied to 21 additional ligands and 19 additional metal fragments, with modest increases in the average error.

In solution, the first four of the five hidden descriptors obtained in our statistical analysis can be associated to qualitative concepts usually associated to metal-ligand bond: σ-donation, π interactions, *cis* influence and covalent character. This is an encouraging result because it recovers, from a purely statistical analysis, concepts that are well entrenched in the conventional understanding of metal-ligand bonding. Our model goes beyond the confirmation of qualitative concepts, as it provides a unified framework for the quantitative definition of scales[75] for each of these properties for ligands as diverse as oxo and dihydrogen, and it opens the possibility of evaluating at the same level the properties of metal fragments, which are far less intuitive.

The statistical analysis for the identification of hidden descriptors reported in this work is not necessarily limited to the analysis of metal-ligand bonding, as it can be applied to a variety of other problems in chemistry. Some of these applications are currently under way in our research laboratory.

Acknowledgement

The authors thank the CERCA Programme (Generalitat de Catalunya) and MINECO (project CTQ2017-87792-R) for financial support. OL thanks Repsol and ONCE foundations for his postdoctoral fellowship.

Supporting Information Available

Multiplicity and geometrical structure associated to each metal fragment. BDE matrices, hidden descriptor tables, errors by number of singular values and diagonal matrix **W** for the initial and actual training sets, as well as the extended and full sets. Description of the procedure used to select the actual training set. Procedure for the estimation of BDEs of new ligands or metal fragments and errors in the estimation of BDEs for the full data set. Details of the study of the role of dispersion and water solvation. List of conventional descriptors and best correlations between them and the hidden descriptors in vacuum and water. Correlation between hidden descriptors and BDEs. Octave scripts used for the statistical treatments.

References

(1) Chatt, J.; Duncanson, L. A. Olefin coordination compounds. Part III. Infra-red spectra and structure: attempted preparation of acetylene complexes. *J. Chem. Soc.* **1953**, 2939–2947.

(2) Dewar, M. J. S. A review of π complex theory. *Bull. Chim. Soc. Fr.* **1951**, C71–C79.

(3) Rusanova, J.; Rusanov, E.; Gorelski, S. I.; Christendat, D.; Popescu, R.; Farah, A. A.; Beaulac, R.; Reber, C.; Lever, A. B. P. The very covalent diammino(o-benzoquinonediimine) dichlororuthenium(II). An example of very strong pi-back-donation. *Inorg. Chem.* **2006**, *45*, 6246–6262.

(4) Bistoni, G.; Belpassi, L.; Tarantelli, F. Disentanglement of Donation and Back-Donation Effects on Experimental Observables: A Case Study of Gold-Ethyne Complexes. *Angew. Chem. Int. Ed.* **2013**, *52*, 11599–11602.

(5) Ahlquist, M. S. G.; Norrby, P. O. Dispersion and Back-Donation Gives Tetracoordinate $[Pd(PPh_3)_4]$. *Angew. Chem. Int. Ed.* **2011**, *50*, 11794–11797.

(6) Triguero, L.; Föhlisch, A.; Väterlein, P.; Hasselströlm, J.; Weinelt, M.; Pettersson, L. G. M.; Luo, Y.; Ågren, H.; Nilsson, A. Direct experimental measurement of donation/back-donation in unsaturated hydrocarbon bonding to metals. *J. Am. Chem. Soc* **2000**, *122*, 12310–12316.

(7) Clot, E.; Besora, M.; Maseras, F.; Megret, C.; Eisenstein, O.; Oelckers, B.; Perutz, R. N. Bond energy M-C/H-C correlations: dual theoretical and experimental approach to the sensitivity of M-C bond strength to substituents. *Chem. Commun.* **2003**, 490–491.

(8) Clot, E.; Megret, C.; Eisenstein, O.; Perutz, R. N. Exceptional Sensitivity of Metal-Aryl Bond Energies to ortho-Fluorine Substituents: Influence of the Metal, the Coordination Sphere, and the Spectator Ligands on M-C/H-C Bond Energy Correlations. *J. Am. Chem. Soc.* **2009**, *131*, 7817–7827.

(9) Guihaume, J.; Clot, E.; Eisenstein, O.; Perutz, R. N. Importance of palladium-carbon bond energies in direct arylation of polyfluorinated benzenes. *Dalton Trans.* **2010**, *39*, 10510–10519.

(10) Devarajan, D.; Gunnoe, T. B.; Ess, D. H. Theory of Late-Transition-Metal Alkyl and Heteroatom Bonding: Analysis of Pt, Ru, Ir, and Rh Complexes. *Inorg. Chem.* **2012**, *51*, 6710–6718.

(11) Wodrich, M. D.; Busch, M.; Corminboeuf, C. Accessing and predicting the kinetic profiles of homogeneous catalysts from volcano plots. *Chem. Sci.* **2016**, *7*, 5723–5735.

(12) Wang, Y.; Montoya, J. H.; Tsai, C.; Ahlquist, M. S. G.; Norskov, J. K.; Studt, F. Scaling Relationships for Binding Energies of Transition Metal Complexes. *Catal. Lett.* **2016**, *146*, 304–308.

(13) Tolman, C. A. Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalyis. *Chem. Rev.* **1977**, *77*, 313–348.

(14) Perrin, L.; Clot, E.; Eisenstein, O.; Loch, J.; Crabtree, R. H. Computed Ligand Electronic Parameters from Quantum Chemistry and Their Relation to Tolman Parameters, Lever Parameters, and Hammett Constants. *Inorg. Chem.* **2001**, *40*, 5806–5811.

(15) Ciancaleoni, G.; Scafuri, N.; Bistoni, G.; Macchioni, A.; Tarantelli, F.; Zuccaccia, D.; Belpassi, L. When the Tolman Electronic Parameter Fails: A Comparative DFT and Charge Displacement Study of $[(L)Ni(CO)_3]^{0/-}$ and $[(L)Au(CO)^{0/+}$. *Inorg. Chem.* **2014**, *53*, 9907–9916.

(16) Kalescky, R.; Kraka, E.; Cremer, D. New Approach to Tolman's Electronic Parameter Based on Local Vibrational Modes. *Inorg. Chem.* **2014**, *53*, 478–495.

(17) Fey, N. Lost in chemical space? Maps to support organometallic catalysis. *Chem. Cent. J.* **2015**, *9*, 38.

(18) Fey, N. The contribution of computational studies to organometallic catalysis: descriptors, mechanisms and models. *Dalton Trans.* **2010**, *39*, 296–310.

(19) Gusev, D. G. Electronic and Steric Parameters of 76 N-Heterocyclic Carbenes in $Ni(CO)_3(NHC)$. *Organometallics* **2009**, *28*, 6458–6461.

(20) Nelson, D. J.; Nolan, S. P. Quantifying and understanding the electronic properties of N-heterocyclic carbenes. *Chem. Soc. Rev.* **2013**, *42*, 6723–6753.

(21) Liske, A.; Verlinden, K.; Buhl, H.; Schaper, K.; Ganter, C. Determining the pi-Acceptor Properties of N-Heterocyclic Carbenes by Measuring the Se-77 NMR Chemical Shifts of Their Selenium Adducts. *Organometallics* **2013**, *32*, 5269–5272.

(22) Dorta, R.; Stevens, E. D.; Scott, N. M.; Costabile, C.; Cavallo, L.; Hoff, C. D.; Nolan, S. P. Steric and electronic properties of N-heterocyclic carbenes (NHC): A detailed study on their interaction with $Ni(CO)_4$. *J. Am. Chem. Soc.* **2005**, *127*, 2485–2495.

(23) Falivene, L.; Credendino, R.; Poater, A.; Petta, A.; Serra, L.; Oliva, R.; Scarano, V.; Cavallo, L. SambVca 2. A Web Tool for Analyzing Catalytic Pockets with Topographic Steric Maps. *Organometallics* **2016**, *35*, 2286–2293.

(24) Droege, T.; Glorius, F. The Measure of All Rings-N-Heterocyclic Carbenes. *Angew. Chem. Int. Ed.* **2010**, *49*, 6940–6952.

(25) Pickup, O. J. S.; Khazal, I.; Smith, E. J.; Whitwood, A. C.; Lynam, J. M.; Bolaky, K.; King, T. C.; Rawe, B. W.; Fey, N. Computational Discovery of Stable Transition-Metal Vinylidene Complexes. *Organometallics* **2014**, *33*, 1751–1761.

(26) Ciancaleoni, G.; Biasiolo, L.; Bistoni, G.; Macchioni, A.; Tarantelli, F.; Zuccaccia, D.; Belpassi, L. Selectively Measuring pi Back-Donation in Gold(I) Complexes by NMR Spectroscopy. *Chem. Eur. J.* **2015**, *21*, 2467–2473.

(27) Leyssens, T.; Peeters, D.; Orpen, A. G.; Harvey, J. N. How important is metal-ligand back-bonding toward YX_3 ligands (Y = N, P, C, Si)? an NBO analysis. *Organometallics* **2007**, *26*, 2637–2645.

(28) Suresh, C.; Koga, N. Quantifying the electronic effect of substituted phosphine ligands via molecular electrostatic potential. *Inorg. Chem.* **2002**, *41*, 1573–1578.

(29) Urbano-Cuadrado, M.; Carbo, J. J.; Maldonado, A. G.; Bo, C. New quantum mechanics-based three-dimensional molecular Descriptors for use in QSSR approaches: Application to asymmetric catalysis. *J. Chem. Inf. Model.* **2007**, *47*, 2228–2234.

(30) Aguado-Ullate, S.; Guasch, L.; Urbano-Cuadrado, M.; Bo, C.; Carbo, J. J. 3D-QSPR models for predicting the enantioselectivity and the activity for asymmetric hydroformylation of styrene catalyzed by Rh-diphosphane. *Catal. Sci. Technol.* **2012**, *2*, 1694–1704.

(31) Aguado-Ullate, S.; Urbano-Cuadrado, M.; Villalba, I.; Pires, E.; Garcia, J. I.; Bo, C.; Carbo, J. J. Predicting the Enantioselectivity of the Copper-Catalysed Cyclopropanation of Alkenes by Using Quantitative Quadrant-Diagram Representations of the Catalysts. *Chem. Eur. J.* **2012**, *18*, 14026–14036.

(32) Garcia-Lopez, D.; Cid, J.; Marques, R.; Fernandez, E.; Carbo, J. J. Quantitative Structure-Activity Relationships for the Nucleophilicity of Trivalent Boron Compounds. *Chem. Eur. J.* **2017**, *23*, 5066–5075.

(33) Angermund, K.; Baumann, W.; Dinjus, E.; Fornika, R.; Gorls, H.; Kessler, M.; Kruger, C.; Leitner, W.; Lutz, F. Complexes [(P-2)Rh(hfacac)] as model compounds for the fragment [(P-2)Rh] and as highly active catalysts for CO2 hydrogenation: The accessible molecular surface (AMS) model as an approach to quantifying the intrinsic steric properties of chelating ligands in homogeneous catalysis. *Chem. Eur. J.* **1997**, *3*, 755–764.

(34) Lipkowitz, K.; D'Hue, C.; Sakamoto, T.; Stack, J. Stereocartography: A computational mapping technique that can locate regions of maximum stereoinduction around chiral catalysts. *J. Am. Chem. Soc.* **2002**, *124*, 14255–14267.

(35) Poater, A.; Ragone, F.; Mariz, R.; Dorta, R.; Cavallo, L. Comparing the Enantioselective Power of Steric and Electrostatic Effects in Transition-Metal-Catalyzed Asymmetric Synthesis. *Chem. Eur. J.* **2010**, *16*, 14348–14353.

(36) Foscato, M.; Venkatraman, V.; Occhipinti, G.; Alsberg, B. K.; Jensen, V. R. Automated Building of Organometallic Complexes from 3D Fragments. *J. Chem. Inf. Model.* **2014**, *54*, 1919–1931.

(37) Chu, Y.; Heyndrickx, W.; Occhipinti, G.; Jensen, V. R.; Alsberg, B. K. An Evolutionary Algorithm for de Novo Optimization of Functional Transition Metal Compounds. *J. Am. Chem. Soc.* **2012**, *134*, 8885–8895.

(38) Foscato, M.; Occhipinti, G.; Venkatraman, V.; Alsberg, B. K.; Jensen, V. R. Automated Design of Realistic Organometallic Molecules from Fragments. *J. Chem. Inf. Model.* **2014**, *54*, 767–780.

(39) Foscato, M.; Houghton, B. J.; Occhipinti, G.; Deeth, R. J.; Jensen, V. R. Ring Closure To Form Metal Chelates in 3D Fragment-Based de Novo Design. *J. Chem. Inf. Model.* **2015**, *55*, 1844–1856.

(40) Venkatraman, V.; Gupta, M.; Foscato, M.; Svendsen, H. F.; Jensen, V. R.; Alsberg, B. K. Computer-aided molecular design of imidazole-based absorbents for CO_2 capture. *Int. J. Greenh. Gas Con.* **2016**, *49*, 55–63.

(41) Maldonado, A. G.; Rothenberg, G. Predictive modeling in homogeneous catalysis: a tutorial. *Chem. Soc. Rev.* **2010**, *39*, 1891–1902.

(42) Ras, E.-J.; Mckay, B.; Rothenberg, G. Understanding Catalytic Biomass Conversion Through Data Mining. *Top. Catal.* **2010**, *53*, 1202–1208.

(43) Westerhuis, J. A.; Hageman, J. A.; Fruhauf, H.-W.; Rothenberg, G. Understanding ligand diversity. *Chim. Oggi Chem. Today* **2007**, *25*, 28–32.

(44) Landman, I. R.; Paulson, E. R.; Rheingold, A. L.; Grotjahn, D. B.; Rothenberg, G. Designing bifunctional alkene isomerization catalysts using predictive modelling. *Catal. Sci. Technol.* **2017**, *7*, 4842–4851.

(45) Piou, T.; Romanov-Michailidis, F.; Romanova-Michaelides, M.; Jackson, K. E.; Semakul, N.; Taggart, T. D.; Newell, B. S.; Rithner, C. D.; Paton, R. S.; Rovis, T. Correlating Reactivity and Selectivity to Cyclopentadienyl Ligand Properties in Rh(III)-Catalyzed C-H Activation Reactions: An Experimental and Computational Study. *J. Am. Chem. Soc.* **2017**, *139*, 1296–1310.

(46) Ardkhean, R.; Mortimore, M.; Paton, R. S.; Fletcher, S. P. Formation of quaternary centres by copper catalysed asymmetric conjugate addition to beta-substituted cyclopentenones with the aid of a quantitative structure-selectivity relationship. *Chem. Sci.* **2018**, *9*, 2628–2632.

(47) Harper, K. C.; Sigman, M. S. Three-Dimensional Correlation of Steric and Electronic Free Energy Relationships Guides Asymmetric Propargylation. *Science* **2011**, *333*, 1875–1878.

(48) Harper, K. C.; Bess, E. N.; Sigman, M. S. Multidimensional steric parameters in the analysis of asymmetric catalytic reactions. *Nat. Chem.* **2012**, *4*, 366–374.

(49) Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A. The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016**, *49*, 1292–1301.

(50) Niemeyer, Z. L.; Milo, A.; Hickey, D. P.; Sigman, M. S. Parameterization of phosphine ligands reveals mechanistic pathways and predicts reaction outcomes. *Nat. Chem.* **2016**, *8*, 611–618.

(51) Guo, J. Y.; Minko, Y.; Santiago, C. B.; Sigman, M. S. Developing Comprehensive Computational Parameter Sets To Describe the Performance of Pyridine-Oxazoline and Related Ligands. *Acs Catalysis* **2017**, *7*, 4144–4151.

(52) Santiago, C. B.; Guo, J. Y.; Sigman, M. S. Predictive and mechanistic multivariate linear regression models for reaction development. *Chem. Sci.* **2018**, *9*, 2398–2412.

(53) Fey, N.; Haddow, M. F.; Harvey, J. N.; McMullin, C. L.; Orpen, A. G. A ligand knowledge base for carbenes (LKB-C): maps of ligand space. *Dalton Trans.* **2009**, 8183–8196.

(54) Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Murray, P.; Orpen, A. G.; Osborne, R.; Purdie, M. Computational descriptors for chelating P,P- and P,N-donor ligands. *Organometallics* **2008**, *27*, 1372–1383.

(55) Qu, X.; Latino, D. A. R. S.; Aires-de Sousa, J. A big data approach to the ultra-fast prediction of DFT-calculated bond energies. *J. Cheminform.* **2013**, *5*, 34.

(56) von Lilienfeld, O. A. Quantum Machine Learning in Chemical Compound Space. *Angew. Chem. Int. Ed.* **2018**, *57*, 4164–4169.

(57) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(58) Carleo, G.; Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **2017**, *355*, 602–605.

(59) Dral, P. O.; Owens, A.; Yurchenko, S. N.; Thiel, W. Structure-based sampling and self-correcting machine learning for accurate calculations of potential energy surfaces and vibrational levels. *J. Chem. Phys.* **2017**, *146*, 244108.

(60) Janet, J. P.; Kulik, H. J. Predicting electronic structure properties of transition metal complexes with neural networks. *Chem. Sci.* **2017**, *8*, 5137–5152.

(61) Yada, A.; Nagata, K.; Ando, Y.; Matsumura, T.; Ichinoseki, S.; Sato, K. Machine Learning Approach for Prediction of Reaction Yield with Simulated Catalyst Parameters. *Chem. Lett.* **2018**, *47*, 284–287.

(62) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Muller, K. R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419.

(63) Frisch, M. J. et al. Gaussian 09. 2013.

(64) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5653.

(65) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(66) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parameterization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*.

(67) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row atoms. *J. Chem. Phys.* **1988**, *89*, 2193–218.

(68) Petersson, G. A.; Al-Laham, M. A. A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *J. Chem. Phys.* **1991**, *94*, 6081–6090.

(69) Alvarez-Moreno, M.; de Graaf, C.; Lopez, N.; Maseras, F.; Poblet, J. M.; Bo, C. Managing the Computational Chemistry Big Data Problem: The ioChem-BD Platform. *J. Chem. Inf. Model.* **2015**, *55*, 95–103.

(70) Eaton, J. W.; Bateman, D.; Hauberg, S. *GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations*; CreateSpace Independent Publishing Platform, 2009; ISBN 1441413006.

(71) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3093.

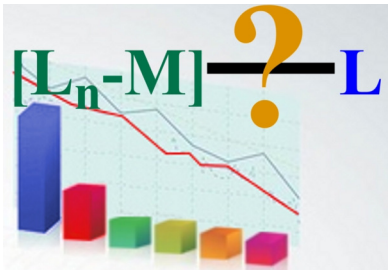
(72) Scalmani, G.; Frisch, M. J. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **2010**, *132*, 114110.

(73) Deprettere, F., *SVD and Signal Processing: Algorithms, Analysis and Applications.*; Elsevier Science Publishers, 1988.

(74) Stone, J. V., *Independent Component Analysis: A Tutorial Introduction*; MIT Press, 2004.

(75) Olmos, A.; Gava, R.; Noverges, B.; Bellezza, D.; Jacob, K.; Besora, M.; Sameera, W. M. C.; Etienne, M.; Maseras, F.; Asensio, G.; Caballero, A.; Pérez, P. J. Measuring the Relative Reactivity of the Carbon–Hydrogen Bonds of Alkanes as Nucleophiles. *Angew. Chem. Int. Ed.* **2018**, *57*, 13848–13852.

Graphical TOC Entry



Synopsis

Crunching numbers to understand: The application of statistical techniques to DFT-generated values for bond dissociation energies leads to hidden descriptors and a better understanding of the underlying contributions to the metal-ligand bond.